

Interpretation of QSAR Models: Mining Structural Patterns Taking into Account Molecular Context

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim The study focused on QSAR model interpretation. The goal was to develop a workflow for the identification of molecular fragments in different contexts important for the property modelled. Using a previously established approach – Structural and physicochemical interpretation of QSAR models (SPCI) – fragment contributions were calculated and their relative influence on the compounds' properties characterised. Analysis of the distributions of these contributions using Gaussian mixture modelling was performed to identify groups of compounds (clusters) comprising the same fragment, where these fragments had substantially different contributions to the property studied. SMARTSminer was used to detect patterns discriminating groups of compounds from each other and visual inspection if the former did not help. The approach was applied to analyse the toxicity, in terms of 40 hour inhibition of growth, of 1984 compounds to *Tetrahymena pyriformis*. The results showed that the clustering technique correctly identified known toxicophoric patterns: it detected groups of compounds where fragments have specific molecular context making them contribute substantially more to toxicity. The results show the applicability of the interpretation of QSAR models to retrieve reasonable patterns, even from data sets consisting of compounds having different mechanisms of action, something which is difficult to achieve using conventional pattern/data mining approaches.

<http://dx.doi.org/10.1002/minf.201800084>

Keywords

Gaussian Mixture Modeling, pattern mining, QSAR interpretation

References

- [1] P. Polishchuk, J. Chem. Inf. Model. 2017, 57, 2618–2639
- [2] OECD Papers 2006, 6, 79–157
- [3] A. G. Leach, H. D. Jones, D. A. Cosgrove, P. W. Kenny, L. Ruston, P. MacFaul, J. M. Wood, N. Colclough, B. Law, J. Med. Chem. 2006, 49, 6672–82
- [4] S. Riniker, G. A. Landrum, J. Cheminf. 2013, 5, 43
- [5] P. G. Polishchuk, V. E. Kuz'min, A. G. Artemenko, E. N. Muratov, Mol. Inf. 2013, 32, 843–853
- [6] Y. Sushko, S. Novotarskyi, R. Korner, J. Vogt, A. Abdelaziz, I. V. Tetko, J. Cheminf. 2014, 6, 48
- [7] V. E. Kuz'min, A. G. Artemenko, E. N. Muratov, J. Comput.-Aided Mol. Des. 2008, 22, 403–421
- [8] V. E. Kuz'min, A. G. Artemenko, P. G. Polishchuk, E. N. Muratov, A. I. Khromov, A. V. Liahovskiy, S. A. Andronati, S. Y. Makan, J. Mol. Model. 2005, 11, 457–467

- [9] P. Polishchuk, O. Tinkov, T. Khristova, L. Ognichenko, A. Kosinskaya, A. Varnek, V. Kuz'min, *Chem. Inf. Model.* 2016, 56, 1455–1469
- [10] Y. Sushko, S. Novotarskyi, R. Korner, J. Vogt, A. Abdelaziz, I. Tetko, *J. Cheminf.* 2014, 6, 48
- [11] Y.-Y. Zhang, H. Liu, S. G. Summerfield, C. N. Luscombe, J. Sahi, *Mol. Pharmaceutics* 2016
- [12] G. Papadatos, M. Alkarouri, V. J. Gillet, P. Willett, V. Kadirkamanathan, C. N. Luscombe, G. Bravi, N. J. Richmond, S. D. Pickett, J. Hussain, J. M. Pritchard, A. W. J. Cooper, S. J. F. Macdonald, *J. Chem. Inf. Model.* 2010, 50, 1872–1886
- [13] S. Bietz, K. T. Schomburg, M. Hilbig, M. Rarey, *J. Chem. Inf. Model.* 2015, 55, 1535–46
- [14] V. Ruusmann, U. Maran, *J. Comput.-Aided Mol. Des.* 2013, 27, 583–603
- [15] Standardizer 16.9.12., 16.9.12.; ChemAxon (<http://www.chemaxon.com>): 2016
- [16] P. G. Polishchuk SPCI: Structural and physico-chemical interpretation tool, <https://github.com/DrrDom/spci>
- [17] cxcalc 16.9.12, 16.9.12; ChemAxon (<http://www.chemaxon.com>): 2016
- [18] P. G. Polishchuk Simplex representation of molecular structure – a chemoinformatic tool for calculation of simplex descriptors v. 1.1.1. <https://github.com/DrrDom/sirms>. <https://github.com/DrrDom/sirms>
- [19] Scikit-learn 0.18, 0.18; Pedregosa et al.: 2016
- [20] RDKit, Open-Source Cheminformatics 2017.09.1 <http://www.rdkit.org>
- [21] C. Fraley, A. E. Raftery, *J. Am. Stat. Assoc.* 2002, 97, 611–631
- [22] T. B. M. Adrian E. Raftery, Luca Scrucca, mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report No. 597; 597; University of Washington: 2012
- [23] P. G. Polishchuk Analysis of fragments contributions calculated by SPCI software. <https://github.com/DrrDom/rspci>
- [24] M. T. Cronin, *Environ. Sci. Process. Impacts* 2017, 19, 213–220
- [25] T. W. Schultz, M. T. Cronin, *J. Chem. Inf. Comput. Sci.* 1999, 39, 304–9
- [26] M. T. Cronin, J. C. Dearden, J. C. Duffy, R. Edwards, N. Manga, A. P. Worth, A. D. Worgan, *SAR QSAR Environ. Res.* 2002, 13, 167–76
- [27] M. T. D. Cronin, Quantitative Structure-Activity Relationships (QSARs) – Applications and Methodology; In *Recent Advances in QSAR Studies*, T. Puzyn, J. L. M. Cronin, Ed. Springer: London, 2010; pp 3–11
- [28] C. M. Ellison, P. Piechota, J. C. Madden, S. J. Enoch, M. T. Cronin, *Environ. Sci. Technol.* 2016, 50, 3995–4007
- [29] H. J. M. Verhaar, C. J. van Leeuwen, J. L. M. Hermens, *Chemosphere* 1992, 25, 471–491
- [30] C. M. Ellison, J. C. Madden, M. T. Cronin, S. J. Enoch, *Chemosphere* 2015, 139, 146–54
- [31] T. Fujita, J. Isawa, C. Hansch, *J. Am. Chem. Soc.* 1964, 86, 5175–5180
- [32] S. J. Enoch, C. M. Ellison, T. W. Schultz, M. T. Cronin, *Crit. Rev. Toxicol.* 2011, 41, 783–802
- [33] C. M. Ellison, M. T. Cronin, J. C. Madden, T. W. Schultz, *SAR QSAR Environ. Res.* 2008, 19, 751–83
- [34] T. W. Schultz, M. T. Cronin, T. I. Netzeva, A. O. Aptula, *Chem. Res. Toxicol.* 2002, 15, 1602–9
- [35] Y. K. Koleva, J. C. Madden, M. T. Cronin, *Chem. Res. Toxicol.* 2008, 21, 2300–12
- [36] T. W. Schultz, T. I. Netzeva, D. W. Roberts, M. T. Cronin, *Chem. Res. Toxicol.* 2005, 18, 330–41
- [37] J. S. Jaworska, R. S. Hunter, T. W. Schultz, *Arch. Environ. Contam. Toxicol.* 1995, 29, 86–93
- [38] M. T. Cronin, S. E. Bryant, J. C. Dearden, T. W. Schultz, *SAR QSAR Environ. Res.* 1995, 3, 1–13
- [39] R. Jaffe, *Environ. Pollut.* 1991, 69, 237–57
- [40] A. O. Aptula, D. W. Roberts, M. T. Cronin, T. W. Schultz, *Chem. Res. Toxicol.* 2005, 18, 844–54
- [41] F. Bajot, M. T. Cronin, D. W. Roberts, T. W. Schultz, *SAR QSAR Environ. Res.* 2011, 22, 51–65
- [42] S. J. Enoch, M. T. Cronin, T. W. Schultz, J. C. Madden, *Chemosphere* 2008, 71, 1225–32